

EFFECT OF A REVIEW PROCEDURE ON FINAL EXAMINATION
PERFORMANCE IN A MODIFIED PSI COURSE

An abstract of a Thesis by
Ruth M. Hurst
September 1979
Drake University
Advisor: Larry A. Alferink

The problem. PSI component analyses have virtually overlooked the effect of a review component. In addition, experimental design for PSI research has typically involved group studies. The present study investigated the use of a review procedure in a modified PSI course and the effect it had on final examination performance utilizing a within subject design.

Procedure. Five hundred forty-two Psychology 1 students who completed the final exam were exposed to a review procedure. The procedure required students to answer questions for credit from 3 review chapters when taking other chapter tests during the course. Independent raters were asked to rate final exam questions in terms of difficulty and to answer each of the questions. The final exam included items that had been unavailable to students prior to the exam and items which had been available prior to the exam.

Findings. Final exam performance was unchanging from early chapters to late chapters while chapter mastery increased from early chapters to late chapters. Raters' ratings and performance covaried inversely across chapters. Raters found early chapters to be similar in difficulty to later chapters although performance was lower on later chapters than earlier chapters. Novel questions introduced variability into final exam performance of the students.

Conclusions. Final exam performance on early chapters which included the 3 review chapters was at a level higher than would have been expected had the review procedure not been utilized. Raters' data indicated that this was not due to "easy" questions from early chapters being on the final. When students' performance on novel questions was considered, the effect of the review procedure was more dramatic. The within subject design employed was a viable research design in demonstrating the effect of the review procedure.

Recommendations. In order to enhance performance on final examinations, a procedure should be used which would require review of all chapters or units of material covered in a course. Variables such as question novelty and difficulty should be considered in future research as these variables may effect student performance.

EFFECT OF A REVIEW PROCEDURE ON FINAL EXAMINATION
PERFORMANCE IN A MODIFIED PSI COURSE

A Thesis
Presented to
The School of Graduate Studies
Drake University

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts

by
Ruth M. Hurst
September 1979

EFFECT OF A REVIEW PROCEDURE ON FINAL EXAMINATION
PERFORMANCE IN A MODIFIED PSI COURSE

by

Ruth M. Hurst

Approved by Committee:

Larry A. Alford
Chairperson

Phillip K. Duncan

Russell F. Marshall

Earle L. Canfield
Dean of the School of Graduate Studies

TABLE OF CONTENTS

	PAGE
INTRODUCTION	1
METHOD	4
RESULTS	9
DISCUSSION	19
REFERENCES	24

LIST OF TABLES

TABLE	PAGE
1. Point bi-serial correlations of raters' answers and ratings	15
2. Distribution across ratings by chapter	16

LIST OF FIGURES

FIGURE	PAGE
1. Comparison of final examination performance and chapter test mastery	10
2. Percent correct and difficulty ratings for independent raters	13
3. Comparison of final examination performance on novel and non-novel items	17

CHAPTER I

INTRODUCTION

In a 1968 description of the Keller Plan, also known as Personalized Systems of Instruction (PSI), Keller included the following as components of PSI: 1) student self-pacing of progress through the course, 2) a unit mastery requirement for proceeding to the next unit, 3) lectures and demonstrations as motivational devices, 4) written materials, including study guides to guide student learning, and 5) undergraduate teaching assistants to act as proctors who were to aid in facilitating repeated testing, immediate test results, and tutoring. It is not clear from Keller's (1968) description or from other descriptions of the Keller Plan (Hursh, 1976; Kulik, Kulik, & Carmichael, 1974) whether the component involving unit mastery functionally included a review procedure; i.e., whether students would be tested over previous units on later unit tests or whether mastery of early units was prerequisite for mastery of later units.

Of the many component analyses of PSI courses at the university level (e.g., Hursh, 1976; Johnson & Ruskin, 1977; Kulik et al., 1974; Robin, 1976), few have specifically addressed the question of the efficacy of a review procedure. Since it is important that students master material presented in any course, the investigation of a review

procedure which may enhance mastery of material as measured by examination performance can be considered an important area of study.

Davis (1975), in a study that attempted to analyze a review procedure, reported the use of a group design to show differential performance on final examinations based on whether performance on review items was considered in the grading of unit tests throughout the course; i.e., some students' quiz grades were contingent not only on performance on regular quiz questions but also on performance on review items from previous units. Davis (1975) reported that groups required to answer review items for credit performed significantly better on final and follow-up examinations over the course material than groups not required to answer the review items for credit. Davis (1975) did not indicate specifically whether items on the final examination were novel to the test. Novel items may be defined as test items which did not appear on any chapter test which was administered prior to the final examination or which were not available to students prior to the final exam via the study guide. If the items were not novel, this could have been a factor influencing performance on the final examination. If any of the items on the final had appeared on previous tests, then performance on those items may have been a function of familiarity with the specific items rather than familiarity with the units from which the items were

taken. Further, although Davis (1975) does report a statistically significant difference between groups, a within-subject demonstration of the effectiveness of review procedures would more clearly show that requiring answers to review items on unit tests does produce better performance over reviewed material than unreviewed material when performance is assessed at a later time.

The purpose of this study is to show the effect of a review procedure in a modified PSI course. More specifically, this study will show a within-subject comparison of performance on novel final examination items from units reviewed during the course with performance on novel final examination items from units not reviewed during the course.

CHAPTER II

METHOD

Subjects

Five hundred forty-two Drake University students who enrolled in Introductory Psychology (Psychology 1) and completed the final examination during the Fall 1977 semester served as subjects.

Setting

Classes were held in two auditoriums on the Drake campus. Students were tested throughout the semester in the Psychology Learning Center. The final exam was administered in designated auditoriums on the scheduled final examination dates.

Materials

The text for the course was Psychology: Understanding Behavior (Baron, Byrne, & Kantowitz, 1977). A mastery guide, A Personalized Mastery Guide for Psychology: Understanding Behavior (Santogrossi, 1977) was also available to students. Instructors had an instructor's manual (Leonard) which was unavailable to the students.

A computerized file of multiple choice test items which included items from the available student mastery guide and the unavailable instructor's manual was utilized

in test construction.

Test Construction

A CDC 6400 computer was used to generate copies of chapter tests and copies of the final exam.

Chapter tests. Chapter tests were individually generated 20 item tests. Items were selected for each test in terms of their availability to students prior to the test. Of the 20 test items on each test, approximately (the number varied slightly from test to test) 5 items were from the available guide, approximately 14 items were from the unavailable instructor's manual, and approximately 1 item was from a review chapter. The review item may have been available or unavailable. Review chapters were chapters 1, 3, and 4. A chapter was not considered as a source for review items until that chapter had been completed (chapters were completed sequentially). Chapter 1 was the only chapter with no review items on chapter tests. The additional question for chapter 1 came from unavailable items.

Final examination. The final exam was made up of 100 multiple choice items. There were 9 questions from each of eleven chapters. Of the 9 items, 5 had been available to the students prior to the exam. The remaining 4 items were novel to the exam; i.e., the items were not on any test given prior to the final exam and were not available to the students in the mastery guide.

Procedure

The course was taught using typical PSI procedures with the following modifications: 1) unit mastery was not required before a student was allowed to proceed to the next unit of material, 2) instructor pacing required that students complete chapter tests by certain specified deadlines although students could take tests more rapidly, 3) attendance at a specific number of class activities was required, 4) only the first three attempts on any given chapter were graded for credit although no attempts were graded for credit once mastery was attained, and 5) review items from at least one review chapter (chapters 1, 3, or 4) appeared on each chapter test as specified above.

In order to inform students of the review procedures to be used during the course, a syllabus was distributed to each student during the first class meeting of each section of Psychology 1. After each student received his/her syllabus copy, relevant aspects of the course were presented by the instructor. The following paragraph in the syllabus (p. 3) was referred to by the instructor:

Review items - Most chapter tests, after chapter 1, will include a "review" question as part of the 20 multiple-choice questions on the test. This question will deal with the material in a previous chapter. However, only three chapters will be used for the review items; Chapters 1, 3, and 4.

At the end of the first class meeting, students were encouraged to ask questions about the structure of the

course, including the review procedure. A 10 item test covering syllabus material was taken by each student before he/she was permitted to take chapter tests.

Tests were then taken throughout the semester according to the syllabus schedule and testing procedures. Those students who completed all required chapter tests by the eleventh week of the semester were eligible to take an early final exam at that time. Other students took the final on the regularly scheduled examination date.

Independent raters. Six undergraduate and 2 graduate students who had no experience with the course or the materials used in the course were asked to rate each item which appeared on the final exam in terms of difficulty. In addition, each rater was asked to answer each test item without using reference sources.

Chapter mastery. A random sample of 65 students enrolled in the course was used in computing the percent of students who mastered each chapter during the course.

Review item performance. A random sample of 20 students enrolled in the course was used to assess the percent correct on review items which appeared on chapter tests which had been mastered.

Item analysis. Each item on every final exam completed by students enrolled in the course was evaluated utilizing a TESTAN item analysis.

Experimental Design

A multiply replicated single subject design was used to test the effect of the review procedure. This design allowed for a comparison of review versus non-review chapter performance of students on the final exam. Chapter 2, a non-review chapter, served as a control chapter so a comparison could be made between chapter 2 and the review chapters which all occurred early in the course.

CHAPTER III

RESULTS

A within subject design was used to test the effect of a review procedure on final exam performance in an instructor paced, PSI, introductory psychology course. Students completed final examination items from three review chapters and eight non-review chapters. Independent raters were asked to answer all questions which appeared on the final exam and rate each question on its difficulty. The questions included novel items which had been unavailable to students prior to the final exam and previously available items.

Figure 1 shows the effect the review procedure had on final exam performance by comparing final exam performance to the level of mastery on chapter tests taken during the course. Performance on the final exam was relatively unchanging from early chapters to late chapters with only a slight decreasing trend for chapters 2-7 and an increasing trend for chapters 10-14. The percent of students mastering each chapter for a random sample of 65 students enrolled in the course increased from early chapters (chapters 1-7) to late chapters (chapters 9-14). Chapter 2 had the fewest students mastering it while chapter 14 had the most students mastering it. Final exam performance on

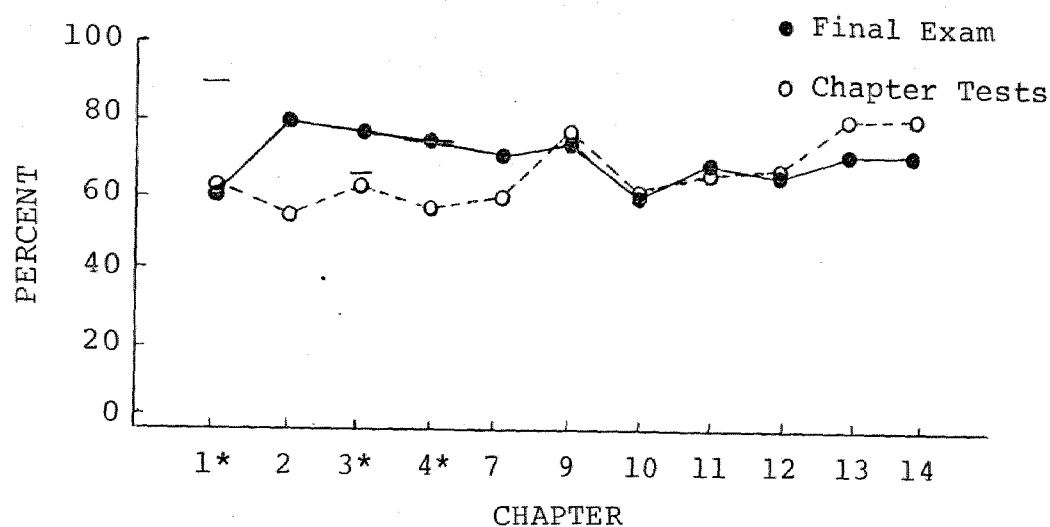


Figure 1. Mean percent of students who took the final exam that correctly answered final exam items by chapter, and percent of students mastering each chapter during the course for a random sample of 65 students enrolled in the course. Horizontal lines over chapters 1, 3, and 4 show the percent correct on items from those chapters which appeared on other mastered chapter tests for a random sample of 20 students enrolled in the course.

*Review chapters

chapters 2-7 was significantly better than the levels of mastery for those chapters while for chapters 1 and 10-14, final exam performance was similar to mastery levels for those chapters.

Final exam performance on chapter 1 yielded a lower mean percent correct than any other chapter on the final exam. Chapter 1 questions included one very difficult novel question which only 18% of the students answered correctly. If this question were excluded from the data analysis, the mean percent correct for chapter 1 would equal 68%. This mean percent correct more closely approximates the level of performance on the other chapters.

Performance on chapter 2, a non-review chapter, was at a level greater than any of the other chapters on the final exam. In addition, fewer students in the random sample of 65 students enrolled in the course mastered chapter 2 than any of the other chapters covered during the course.

Final exam performance on chapter 7, a non-review chapter which occurred relatively early in the course, exceeds the level of mastery for that chapter. The amount by which final exam performance exceeds mastery on chapter 7 closely approximates that by which mastery exceeds performance on chapters 13 and 14.

The percent correct for all review items that appeared on chapter tests is shown in Figure 1 for a random

sample of 20 students enrolled in the course. Percent correct on review items from chapter 1 was 92% correct; from chapter 3, 67% correct; and from chapter 4, 76% correct.

Figure 2 shows a comparison of difficulty ratings and percent correct on items on the final exam for 8 independent raters. Raters rated chapter 1 questions as being slightly more difficult than 7 of the 10 remaining chapters and performed less well on chapter 1 questions than on questions from 4 of the 10 remaining chapters. Raters thus indicated that chapter 1 questions may have been slightly more difficult on the whole than questions from a majority of the other chapters. The independent raters rated chapter 2 questions as being slightly less difficult than questions from 4 other chapters and performed better on chapter 2 questions than on questions from 7 of the 10 remaining chapters. Raters indicated that chapter 2 questions may have been easier than questions from several other chapters.

Raters averaged 51% correct on review chapters and 49% correct on non-review chapters. The average rating for review chapters was 2.8 and for non-review chapters was 3.

Raters performed less well on chapters 11-14 than on earlier chapters while difficulty ratings on chapters 11-14 were slightly higher than ratings on earlier chapters. This suggests that items from later chapters may have been more difficult than items from earlier chapters.

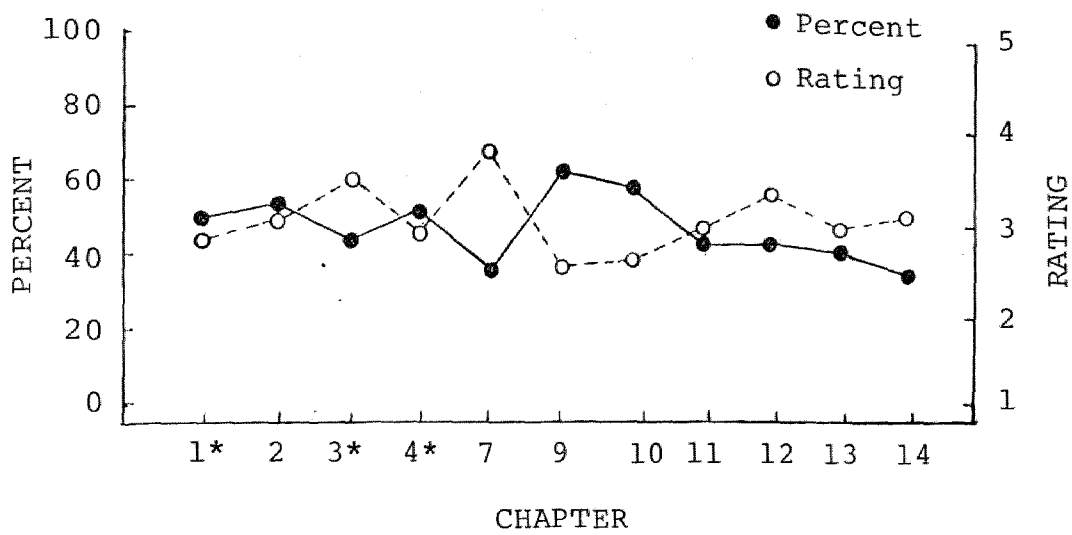


Figure 2. Mean percent correct on final exam items by chapter and mean difficulty rating for final exam items by chapter for 8 independent raters. (Ratings: 1=easy, 5=difficult)

*Review chapters

Raters' ratings and performance on test items varied across chapters with a tendency toward inverse covariation between ratings and performance. Mean point bi-serial correlations shown in Table 1 were all inverse correlations across both chapters and raters. Thus, incorrectly answered items were generally rated as "difficult" and correctly answered items were generally rated as "easy". The mean correlation by raters and by chapters was always a negative correlation between $-.05$ and $-.50$, inclusive.

Table 2 contains the percent of items on the final exam given each of the ratings for each chapter. Raters rated most items as medium in difficulty with fewer items rated at the extremes of the scale.

Figure 3 shows a comparison of novel and non-novel questions for each chapter on the final exam. Performance on novel questions from chapters 2, 3, and 4 was higher than the level of performance on novel questions from other chapters. Novel item performance shows a slight decreasing trend for chapters 7 through 10 and an increasing trend for chapters 10 through 14. The decreasing trend was less marked for early chapters than the increasing trend of the later chapters. Chapters 2, 3, 4, and 14 had the least difference between novel and non-novel question performance. Chapters 1 and 10 had the greatest difference between non-novel and novel question performance. Each chapter had one novel question that few students answered correctly. Thus,

Table 1. Point bi-serial correlation of raters' answers on final exam items to raters' ratings of final exam items in terms of difficulty.

Raters	Chapter											
	1	2	3	4	7	9	10	11	12	13	14	\bar{X}
1	-.53	-.26	-.22	-.04	-.28	0	.16	-.11	-.22	-.27	-.57	-.21
2	-.04	-.07	-.04	-.54	-.47	-.12	-.24	-.17	0	-.43	-.13	-.20
3	-.26	.12	-.07	.21	.14	.28	.14	.31	-.09	-.75	-.64	-.06
4	-.07	-.26	-.14	0	-.39	-.32	-.44	-.19	-.01	-.64	-.23	-.24
5	-.22	-.60	-.10	-.57	-.14	.04	-.56	.30	-.02	-.27	.07	-.19
6	-.32	-.46	-.42	-.18	.13	-.21	-.35	-.46	.28	-.46	-.50	-.27
7	-.49	-.02	-.46	-.33	.31	.09	.11	-.34	.004	-.25	-.47	-.17
8	.09	-.40	-.26	.03	-.61	-.13	-.19	.22	-.23	-.95	-.38	-.26
\bar{X}	-.23	-.24	-.21	-.18	-.16	-.05	-.17	-.05	-.04	-.50	-.36	

Table 2. Percent of items on the final exam given each of 5 ratings by 8 independent raters for each chapter.

Chapter	Rating				
	1	2	3	4	5
1	11	29	35	16	8
2	7	25	35	20	12
3	17	13	29	21	20
4	23	25	23	17	13
7	17	11	29	27	16
9	24	20	32	16	8
10	8	27	40	18	7
11	11	25	42	12	10
12	9	16	31	30	13
13	4	19	35	31	11
14	19	11	30	33	7

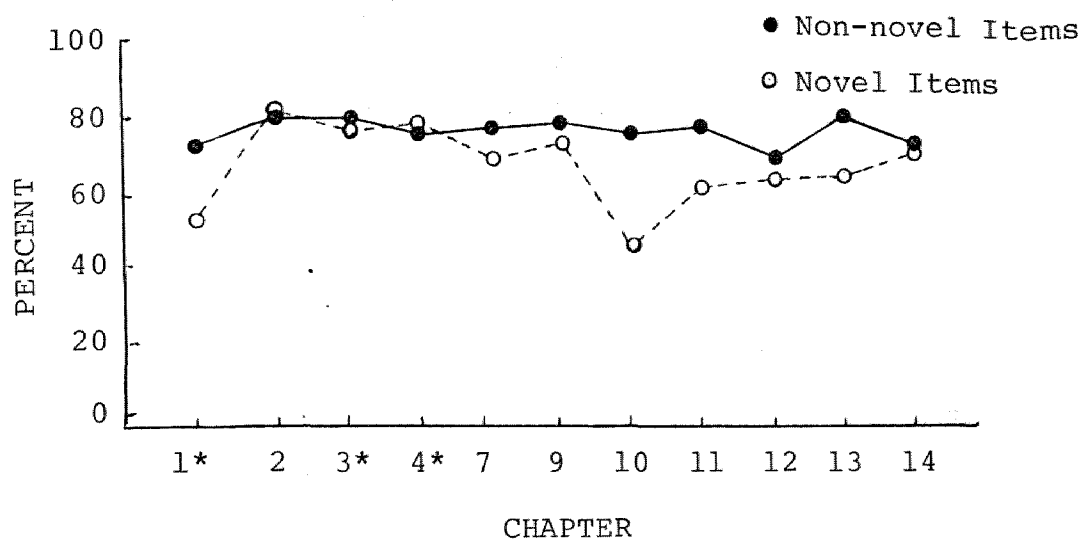


Figure 3. Mean percent of students who took the final exam that correctly answered novel and non-novel items on the final exam by chapter.

*Review chapters

the mean level of performance on novel items on these two chapters was not considered representative and was not considered in the above analysis.

CHAPTER IV

DISCUSSION

Results showed that performance on the final exam was unchanging across chapters. However, during the course, chapter mastery increased from early chapters to late chapters. According to independent raters, the chapters were similar in difficulty. Performance on novel questions was either below or matched performance on non-novel questions appearing on the final exam.

Performance on chapter tests improved during the course from early chapters to late chapters. This may have been due to the practice effects of answering the types of multiple choice questions which appeared on tests. Chapter 2 performance was lower than all other chapters. This chapter was very difficult for most of the students in the course.

Review question data showed that students performed at a higher level on review questions from chapter 1 and lower on review questions from chapters 3 and 4. Students generally answered review questions correctly even though a very weak dependency existed between test mastery and answering review questions correctly; i.e., students could miss up to 3 questions of 20 on a chapters test and still master the test. Review questions accounted for from 0 to 3

questions on any given chapter test. The percent of review items answered correctly from each review chapter was higher than the percent correct on the original exam over that chapter. This suggests that students did review material from review chapters.

Final exam performance was unchanging across chapters. Without the review procedure, final exam performance would probably have more closely approximated mastery performance; i.e., final exam performance would have increased from early chapters to late chapters. Performance on chapters 1 and 10 was lower than performance on other chapters on the final exam due to difficult questions. Excluding the difficult questions from consideration would have made performance on those chapters more closely approximate final exam performance on other chapters. Final exam performance on chapters 2, 3, and 4 was higher than mastery performance for those chapters. For chapter 2, the difference between final exam and mastery performance may show that chapter 2 was functionally a review chapter. The title of chapter 2 was "Biological Bases of Behavior," and many of the concepts and facts covered in that chapter may have been covered in other courses students were enrolled in that semester; i.e., biology, physiology, or nursing courses. Higher performance on chapters 3 and 4 on the final exam when compared to mastery data may have been due to the effect of the review procedure.

Independent raters who rated all questions in terms of difficulty rated chapters to be at about the same difficulty level. Raters also answered the questions they rated. Raters performance varied inversely with the ratings. The slight inverse correlation between correctness of answers and ratings shows that raters may have tended to rate item difficulty depending on whether they thought they answered questions correctly or incorrectly. However, the slight correlations indicate that although this may have been a factor in ratings, it may not have been the only one.

Performance by students on early chapters was slightly better than performance on later chapters. Rater performance was similar to student performance with the exclusion of chapter 10 in that raters did less well on later chapters. However, students performed better on the review chapters relative to the later chapters than did the raters. This suggests that this review procedure may have improved performance on early chapters for students in the course. Other factors could have effected rater performance including the fact that raters answered questions in order, increasing the likelihood that raters may have been more careless in answering the later items. In addition, later chapters may have covered material with which raters were less familiar.

Novel question performance was variable across chapters. Performance on novel questions was higher on

chapters 2, 3, and 4 than on other chapters. Familiar question performance was unchanging across chapters. When considering familiar question performance only, the review procedure looks particularly effective in relation to mastery data characteristics mentioned earlier. Variability between chapters on final exam performance was introduced by novel item performance. Novel questions percent correct approximated non-novel question performance most closely on chapters 2, 3, 4, and 10. Chapter 2 may functionally have been a review chapter, chapters 3 and 4 were review chapters and chapter 14 was the most recent chapter students had experienced in the course. This suggests that recency of chapters and the review procedure probably increased the familiarity of chapters' material and increased the likelihood of students' ability to generalize to unfamiliar (novel) questions.

In summary, the review procedure utilized in this course may have had the effect of improving performance on the final exam when compared to chapter mastery. While these results do not conflict with results suggested by the Davis (1968) research, this study does demonstrate that a within subject design is a viable research design to use in college level courses. An analysis of novel items added to the conclusiveness of this study in that the performance on novel items was differentially effected by the review procedure. The review procedure utilized may be a viable

procedure to use in other similar courses with discrete units of material when improving students' performance on final exams is a desirable outcome.

REFERENCES

- Baron, R., Byrne, D., & Kantowitz, B. Psychology: Understanding behavior. Philadelphia: Saunders, 1977.
- Davis, M. Some effects of an integrated review procedure. In J. M. Johnston (Ed.), Research and technology in college and university teaching. Gainesville: University of Florida, 1975.
- Hursh, D. Personalized systems of instruction: What do the data indicate. Journal of Personalized Instruction, 1976, 1, 91-105.
- Johnson, K., & Ruskin, R. Behavioral instruction: An evaluative review. Washington, D.C.: American Psychological Association, 1977.
- Keller, F. Good-bye, Teacher.... Journal of Applied Behavior Analysis, 1968, 1, 79-89.
- Kulik, J., Kulik, C., & Carmichael, K. The Keller plan in science teaching. Science, 1974, 183, 379-383.
- Leonard, D. Instructor's manual to accompany psychology: Understanding behavior. Philadelphia: Saunders.
- Robin, A. Behavioral instruction in the college classroom. Review of Educational Research, 1976, 46, 313-354.
- Santogrossi, D. Personalized mastery guide for psychology: Understanding behavior. Philadelphia: Saunders, 1977.